AD-A096 413 NORTH CAROLINA UNIV AT CHAPEL HILL DEPT OF STATISTICS F/G 12/1
ROBUST ESTIMATION IN HETEROSCEDASTIC LINEAR MODELS.(U)
FEB 81 R J CARROLL, D RUPPERT AFOSR-80-0080
UNCLASSIFIED AFOSR-TR-81-0219 NL

1 OF 1
AD A
096413

END
DATE
FILMED
4-81
DTIC

ROBUST ESTIMATION IN HETEROSCEDASTIC LINEAR MODELS

Raymond J. Carroll* and David Ruppert**

AMS 1970 subject classifications: Primary 62J05, Secondary 62G35

*Supported by the Air Force Office of Scientific Research under grant [redacted] and by the National Heart, Lung, and Blood Institute
AFOSR-80-0080

**Supported by National Science Foundation Grant MCS78-01240.

Key words and phrases. Weighted M-estimates, weighted least squares, unknown variances, estimated weights, asymptotically normal

Abstract We consider a heteroscedastic linear model in which the variances are a parametric function of the mean responses $\cancel{\{\tau_1, \ldots, \tau_N\}}$ and a parameter θ. We propose robust estimates for the regression parameter β and show that, as long as a reasonable starting estimate of θ is available, our estimates of β are asymptotically equivalent to the natural estimate obtained with known variances. A particular method for estimating θ is proposed and shown by Monte-Carlo to work quite well, especially in power and exponential models for the variances. We also briefly discuss a "feedback" estimate of β.

1. Introduction. We consider the heteroscedastic linear model

(1.1) $$Y_i = \tau_i + \sigma_i \epsilon_i, \qquad \tau_i = x_i \beta, \; i = 1, \ldots, N,$$

where $\{x_i\}$ are $(1 \times p)$ design constants, β is a $(p \times 1)$ regression parameter, $\{\epsilon_i\}$ are independent and identically distributed with mean zero and unknown symmetric distribution function F, and $\{\sigma_i\}$ are scaling constants which express the possible heteroscedasticity. Our primary interest is estimation of and inference about the unknown regression parameter β.

Of course, one could ignore the $\{\sigma_i\}$ and use classical methods such as least squares or M-estimation (Huber (1977)), but such estimates are not efficient. In order to make more efficient inference about β, it is necessary to get information about the $\{\sigma_i\}$. In one approach to the problem, the $\{\sigma_i\}$ are assumed completely unknown, but replication is assumed feasible so that the $\{Y_i\}$ occur in groups of equal variance. Recent results in this direction are due to Fuller and Rao (1978). Their results are complicated and the delicate calculations involved seem to depend very heavily on an assumption of Gaussian errors, i.e., $F = \Phi$, the standard normal distribution function. Such a strong dependence on the Gaussian errors is undesirable from the view of robustness; the assumption of an exact normal error distribution is not always tenable and the resulting estimates tend to be inefficient for distributions with heavier tails than the normal distribution. See Huber (1977) for details and further references.

The second approach to the estimation problem for (1.1) avoids the replication assumption by positing a known form for the error variance, i.e.,

(1.2) $$\sigma_i = H(x_i, \beta, \theta),$$

where θ is $(r \times 1)$ vector of unknown coefficients and H is smooth and known. A model such as (1.2) is behind the tests for homoscedasticity developed by Anscombe (1961), Bickel (1978) and Carroll and Ruppert (1981). Of course, in many real problems we suspect a heteroscedastic model because the dispersion of the residuals increases with the magnitude of the fitted values. Thus, it has become quite common to simplify (1.2) by assuming that σ_i is a function of τ_i or $|\tau_i|$, e.g.,

(1.3a) $\qquad \sigma_i = \sigma(1 + |\tau_i|)^\theta$

(1.3b) $\qquad \sigma_i = \sigma|\tau_i|^\theta$ $\qquad$ (Box and Hill (1974))

(1.3c) $\qquad \sigma_i = \sigma\exp(\theta\tau_i)$ $\qquad$ (Bickel (1978))

(1.3d) $\qquad \sigma_i = \sigma(1 + \theta\tau_i^{\,2})^{1/2}$ $\qquad$ (Jobson and Fuller (1980)).

(See Dent and Hildreth (1977) for other models). Following these examples, we will thus assume that for some known H,

(1.4) $\qquad \sigma_i = \sigma H_*(\tau_i, \theta_*) = H(\tau_i, \theta = (\sigma, \theta_*)).$

Our results can be generalized to the model (1.2), but the statements of results and assumptions then become extremely complicated.

Box and Hill (1974) and Jobson and Fuller (1980) both suggest a form of generalized weighted least squares. One obtains estimates of (θ, β), constructs estimated weights $\hat{\sigma}_i$ and then performs ordinary weighted least squares. Their methods are constructed from a normal error assumption and their efficiency depends on this assumption. The maximum likelihood estimates for θ under the normality assumption have a quadratic influence curve and may be particularly non-robust. As argued above, the recent literature demonstrates some acceptance to the notion that estimators should be robust against departures from normality. One purpose of this article is to provide a set of such robust estimates.

Implicit in the work of Box and Hill (1974) and Jobson and Fuller (1980) is the notion that this problem is $\{\sigma_i\}$ adaptable (Bickel (1980), Wald lectures), i.e., the generalized weighted least squares methods are asymptotically equivalent to the "optimal" weighted least squares estimate one would define if the $\{\sigma_i\}$ were actually known up to a scale factor. Our second major aim is to show that there is a wide class of (robust) estimates of β which are $\{\sigma_i\}$ adaptable for many distribution functions F and models (1.4).

2. A class of weighted robust estimates. Suppose we have estimates of $(\hat{\theta}, \hat{\beta})$ which are $N^{\frac{1}{2}}$-consistent, i.e.,

$$(2.1)\qquad \begin{aligned} N^{\frac{1}{2}}(\hat{\theta} - \theta) &= O_p(1) \\ N^{\frac{1}{2}}(\hat{\beta}_0 - \beta) &= O_p(1). \end{aligned}$$

The existence of such estimates is discussed in the next section. We form the estimated σ_i:

$$(2.2)\qquad \hat{\sigma}_i = H(t_i, \hat{\theta}), \; t_i = x_i\hat{\beta}_0 .$$

If the $\{\sigma_i\}$ were known robustness considerations discussed by Huber (1973, 1977) suggest a general class of weighted M-estimates formed by solving the minimization problem in β

$$(2.3)\qquad \Sigma\rho((Y_i - x_i\beta)/\sigma_i) = \text{minimum}$$

Here ρ is taken to be a convex function. If, for example, $\rho(x) = x^2/2$ we get the "optimal" weighted least squares estimate with known weights. In general, the unknown solution to (2.3) is denoted $\hat{\beta}_{opt}$.

The class of estimates we suggest are very simply generated by substituting $\{\hat{\sigma}_i\}$ into (2.3). Taking derivatives, we suggest solving the equation

$$\sum_{i=1}^{N} (x_i'/\hat{\sigma}_i)\psi((Y_i - x_i\beta)/\hat{\sigma}_i) = o, \tag{2.4}$$

with solution $\hat{\beta}$. Throughout we take ψ to be an odd, continuous function. The (non-robust) generalized weighted least squares estimates suggested by Box and Hill (1974) and Jobson and Fuller (1980) fall under the special case of (2.4) when $\psi(x) = x$; both propose possibilities for $\hat{\beta}_0$ and $\hat{\theta}$ of (2.1). As suggested by the literature, choosing ψ bounded should result in reasonably efficient and robust estimates of β.

Define $d_i = x_i/\sigma_i$ and assume that

$$S_N = N^{-1} \sum_{i=1}^{N} d_i' d_i \rightarrow S \text{ (positive definite)}. \tag{2.5}$$

Then by formal Taylor series arguments the optimal robust weighted estimate $\hat{\beta}_{opt}$ which minimizes (2.4) satisfies

$$N^{\frac{1}{2}}(\hat{\beta}_{opt} - \beta) = N^{-\frac{1}{2}} \sum_{i=1}^{N} S^{-1} d_i \psi(\epsilon_i)/E\psi'(\epsilon_1) + o_p(1) \tag{2.6}$$

$$\xrightarrow{\mathcal{L}} N(o,\ E\psi^2 S^{-1}(E\psi')^{-2}).$$

Our main result concerning adaptation is that when (2.1) holds and hence we have a reasonable estimate of θ, then our estimate $\hat{\beta}$ is asymptotically equivalent to the (unknown) $\hat{\beta}_{opt}$. Formally, we have

Theorem 1. Assume (2.1), (2.5), (2.9) and B1 - B8 below. Then

$$N^{\frac{1}{2}}(\hat{\beta} - \hat{\beta}_{opt}) \xrightarrow{P} o, \tag{2.7}$$

so that

$$N^{\frac{1}{2}}(\hat{\beta} - \beta) \xrightarrow{\mathcal{L}} N(o,\ E\psi^2 S^{-1}(E\psi')^{-2}). \tag{2.8}$$

That $\hat{\beta}$ is robust against outliers in the errors when ψ is bounded can be seen by combining (2.6) - (2.7). The resulting influence curve is strikingly similar to the unweighted case in homoscedastic models.

In stating assumptions and proofs we simplify (1.4) to

$$\sigma_i = \exp(h(\tau_i)\theta), \tag{2.9}$$

where h is a function from R to R^r. The model (2.9) includes models (1.3a) - (1.3c), but it is not strictly necessary for the validity of our results. Our reason for considering only (2.9) in the formal aspects of this section is to avoid making already cumbersome notation needlessly complicated. Generalizations to the model (1.4) require that $H(\cdot,\cdot)$ be smooth.

Here are the assumptions.

B1. ψ odd, F symmetric, $0 < E\psi^2(\varepsilon_1) < \infty$, $E\psi' > 0$

B2. As $r \to 0$, $s \to 0$,

$$E\psi((\varepsilon_1 + r)(1 + s)) = rE\psi' + o(|r| + |s|),$$

B3. There exists C_0 such that for all $\delta < 1$, as $K \to \infty$

$$E\left\{\sup|\psi((\varepsilon_1 + r)(1 + s)) - \psi((\varepsilon_1 + r')(1 + s'))| : |r|, |r'|, |s|, |s'| \le K, |r - r'| \text{ and } |s - s'| \le K\delta\right\} \le C_0\delta$$

B4. As $r, s \to 0$,

$$E(\psi((\varepsilon_1 + r)(1 + s)) - \psi(\varepsilon_1))^2 \to 0$$

B5. $$\lim_{N\to\infty} \sup_{i \le N} (\|x_i\| + \|h(\tau_i)\|)N^{-1/2} = 0$$

B6. $\sup_N (N^{-1} \sum_{i=1}^{N} (||x_i||^2 + ||h(\tau_i)||^2)) < \infty$

B7. The functions $\sigma_i = H(\tau_i, \theta)$ are bounded away from zero.

B8. On an interval I containing all the $\{\tau_i\}$, h is Lipschitz continuous. I may be infinite.

The proof is given in Appendix B. Conditions B1 - B6 are similar to those used by Bickel (1975) in his study of one-step M-estimates in the homoscedastic model. Condition B7 merely insures that we do not have infinite weights and condition B8 assures us that when $\sigma_i = H(\tau_i, \theta) = \exp(h(\tau_i)\theta)$, the function H is sufficiently smooth. Details verifying that (2.8) actually follows from (2.6) - (2.7) and the assumptions of Theorem I are easy to fill in.

3. Estimation of θ. In the previous section we have shown that, except for certain technical conditions, one can construct robust weighted estimates of β as long as one has available estimates of θ and β which satisfy (2.1). Preliminary estimates $\hat{\beta}_0$ satisfying (2.1) are readily available and include (under reasonable assumptions) ordinary least squares estimates and ordinary M-estimates; details of sufficient conditions for this are available from the authors in the form of an earlier version of this paper. Bounded influence regression estimates could also be used; see e.g., Krasker and Welsch (1981). In this section we propose a class of estimates of θ which are robust and satisfy (2.1). There are of course many possible ways to construct such estimates, but our method has the necessary theoretical properties as well as encouraging small sample properties; see the next section for details.

To motivate our estimates, suppose the $\{\tau_i\}$ were known, the $\{\sigma_i\}$ satisfy (1.4) and the density of f is proportional to $\exp(-\rho(x))$, where ρ and $\rho' = \psi$ are as in the previous section. This device is common in robustness studies; see

Huber (1977), Bickel and Doksum (1981) and Carroll (1980) for examples. In this instance the log-likelihood for θ is (up to a constant)

$$\ell(\theta) = -\sum_{i=1}^{N} \log H(\tau_i,\theta) - \sum_{i=1}^{N} \rho((Y_i - \tau_i)/H(\tau_i,\theta)). \tag{3.1}$$

Taking derivatives in θ suggests that we solve

$$0 = \ell'(\theta) = \sum_{i=1}^{N} \left\{ \psi\left(\frac{Y_i - t_i}{H(t_i,\theta)}\right)\left(\frac{Y_i - t_i}{H(t_i,\theta)}\right) - 1 \right\} \frac{\partial}{\partial\theta} H(t_i,\theta)/H(t_i,\theta). \tag{3.2}$$

Because the term in brackets in (3.2) is not bounded and hence would in general lead to an unbounded influence function for the estimated θ and an overall lack of robustness in our estimation procedure, we follow the common device used in the homoscedastic case by Huber (1977) and Bickel and Doksum (1981) of replacing $x\psi(x) - 1$ by a function $\chi(\cdot)$, as well as replacing τ_i by $t_i = x_i\hat{\beta}_0$, thus leading to estimates obtained by solving

$$0 = G_N(\theta) = \sum_{i=1}^{N} \chi\,((Y_i - t_i)/H(t_i,\theta))\, \frac{\partial}{\partial\theta} H(t_i,\theta)/H(t_i,\theta). \tag{3.3}$$

Probably the most common choice of $\chi(\cdot)$ in the homoscedastic case is that for the classical Proposal 2:

$$\chi(y) = \psi^2(y) - \int\psi^2(x)(2\pi)^{-1/2}\exp(-x^2/2)dx.$$

This choice of $\chi(\cdot)$ gives bounded influence to our estimates of θ, and thus might reasonably be preferred in our problem to $y\psi(y) - 1$, just as it is in the homoscedastic case; see also Huber (1977, p. 33) for certain optimality properties of this choice. In the case of the special model (2.9), we have

$$G_N(\theta) = \sum_{i=1}^{N} \chi\left((Y_i - t_i)/\exp(h(t_i)\theta)\right)h(t_i). \tag{3.4}$$

We make the assumptions that $\chi(\cdot)$ is an even function with $\chi(0)<0$, $\chi(\infty)>0$. In the model (1.4), σ is a free parameter defined so that

$$(3.5) \qquad E\chi((Y_1 - \tau_1)/\sigma_1) = 0.$$

In the model (1.3a) we have

$$\theta = \begin{pmatrix} \log\sigma \\ \theta_* \end{pmatrix} \qquad h(\tau) = \begin{pmatrix} 1 \\ 1 + |\tau| \end{pmatrix}^T.$$

In many models (such as (1.3a), (1.3b), and (1.3c) with $\tau_i > 0$) one can show that solutions to the equation $G_N(\theta) = G_N(\sigma,\theta_*) = 0$ exist. But we have been unable to show that the solutions are unique, although in all of our examples unique solutions have been obtained. More importantly, one may not wish to consider all possible values of θ, e.g., in models (1.3a) - (1.3c), one may reasonably wish to restrict $|\theta|\le 1.5$. For these reasons, we suggest the following procedure:

(3.6) minimize $||G_N(\theta)|| = ||G_N(\sigma,\theta_*)||$ on the interval $\theta_* \in J$. If the solution is not unique, choose that one with smallest $||\theta_*||$.

The solution to (3.6) is thus well-defined. In all of our examples when θ is unrestricted, the solutions to (3.3) and (3.6) have coincided. In the examples in which we have restricted θ, (3.6) has always had a unique solution even when (3.3) has not had a solution in the restricted space.

An appealing feature of these estimates is that they are natural generalizations of the classical Huber Proposal 2 for the homoscedastic case.

Theorem 2. Assume (2.5),(2.9), (3.5), B5 - B8. Further assume that $N^{\frac{1}{2}}(\hat{\beta}_0 - \beta) = O_p(1)$. Then under C1 - C5 below, if $\hat{\theta}$ solves (3.6), then

$$(3.7) \qquad \hat{\theta} - \theta = O_p(N^{-\frac{1}{2}}).$$

Here are the conditions:

<u>C1</u> $\quad 0 < E\chi^2(\epsilon_1) < \infty$, and χ is nondecreasing on $[0, \infty)$.

<u>C2</u> $\quad$ As $r, s \to 0$ for $A(\chi) > 0$,

$$E\chi((\epsilon_1 + r)(1 + s)) = A(\chi)s + o(|r| + |s|).$$

<u>C3</u> $\quad$ Condition <u>B3</u> holds for χ.

<u>C4</u> $\quad$ Condition <u>B4</u> holds for χ.

<u>C5</u> $\quad$ If λ_N is the minimal eigenvalue of

$$H_N = N^{-1}\sum_{i=1}^{N} h(\tau_i)^T h(\tau_i),$$

then $\liminf \lambda_N = \lambda_0 > 0$.

The proof is given in Appendix C. The conditions are hardly onerous, being similar to those of Bickel (1975), and with only <u>C5</u> affected by the heteroscedasticity. Further details of implementation are discussed in the next section.

One can also introduce redescending M-estimates by using ψ redescending to zero. No change is needed in Theorem 1, while an estimate for θ can be obtained by doing one or two steps of Newton-Raphson for (3.3) from any estimate satisfying (2.1). Proofs are similar to those given in the appendicies.

4. <u>A Monte-Carlo study</u>. Because Theorem 1 is an asymptotic result, we performed a small Monte-Carlo study to assess the small sample properties of $\hat{\beta}$. The model was simple linear regression

$$(4.1) \qquad Y_i = \beta_0 + \beta_1 C_i + \sigma_i\epsilon_i = \tau_i + \sigma_i\epsilon_i, \quad i = 1, \ldots, N.$$

In the study, the $\{C_i\}$ were equally spaced between -2 and +2, and we chose to study the model (1.3a)

$$\sigma_i = \sigma(1 + \theta|\tau_i|) .$$

The experiments were each repeated two hundred times under the following circumstances:

(a) $N = 21$, $\{\epsilon_i\}$ are standard normal, $\sigma = .25$, $\beta_0 = 2$, $\beta_1 = 1$

(b) $N = 41$, $\{\epsilon_i\}$ are $N(o,1)$ with probability $p = .90$ and $N(o, \text{Var} = 9)$ with $p = .10$, $\sigma = .25$, $\beta_0 = 4$, $\beta_1 = 2$. The $\{\epsilon_i\}$ are said to have a contaminated normal distribution. Such distributions are often used in robustness studies.

We made two choices for ψ. First was $\psi(x) = x$, which yields the usual weighted least squares estimate $\hat{\beta}_L$, and the second was Huber's $\psi(x) = \max(-2.0, \min(x, 2.0))$. This gives a version $\hat{\beta}_R$ of our robust weighted estimates. In constructing $\hat{\sigma}_i$ we defined

(4.2) $$\chi(y) = \psi^2(y) - \int\psi^2(y)\theta(y)dy,$$

where $\theta(\cdot)$ is the standard normal density function. This particular $\chi(\cdot)$ is in standard use for Huber's Proposal 2 in homoscedastic models. The choice $\chi(y) = y\psi(y) - 1$ was not chosen because of unbounded influence functions leading to robustness difficulties.

Both $\hat{\beta}_L$ and $\hat{\beta}_R$ were constructed as follows:

Step (i): Let $\beta_\star$ be the unweighted Huber Proposal 2 estimate ($\theta = o$) with χ given by (4.2) and $\psi(x) = \max(-2, \min(x, 2))$.

Step (ii): Solve (3.5) for $(\sigma_\star, \theta_\star)$ form inverse "weights" $w_i^2 = (1 + |t_i|)^{2\theta\star}$, $t_i = x_i\beta_\star$.

Step (iii): Solve a weighted Huber Proposal 2 by simultaneously solving (2.4) for the desired function ψ and the part of (3.5) given by

(4.3) $$\Sigma\chi((Y_i - x_i\beta)/\sigma w_i) = o.$$

The result is $\hat{\beta}_0$.

Step (iv): Repeat steps (ii) and (iii) to obtain $t_i = x_i\hat{\beta}_0$, $\hat{\theta}$, $\hat{\sigma}$, $\hat{\beta}$.

The algorithm given here was chosen so as to reproduce Huber's Proposal 2 in the homoscedastic case $\theta = o$. Direct application of the results of Section 2 involves only solving (2.4) in Step (iii) and gave results essentially indistinguisable from those reported here. In solving for $(\hat{\beta}, \hat{\sigma})$ we used the subroutine ZXGSN of the IMSL library.

In Table #1 we list part of the results of the study. The values listed are ratios of mean square errors for estimating β_1 in model (4.1), the ratio being with respect to the "optimal" robust method one would use if $w_i^* = (1 + |\tau_i|)^\theta$ were actually known, i.e., solve (2.4) and (4.3) simultanuously with the known weights. The study is fairly small but it does seem to indicate that our robust weighted estimate will work in situations in which heteroscedasticity is suspected.

It is important to note that our estimate has MSE never more than 10% larger than the unknown estimate formed with the correct weights, and seems to do better than unweighted estimates when $\theta \neq o$. Note also the robustness feature; the efficiency of the weighted least squares estimates (even the "optimal" one) depends heavily on the normality assumption and is not very high in the contaminated case. All of these results tend to support the applicability of Theorem 1.

We repeated the experiment but with the model

$$\sigma_i = \sigma \exp(\theta |\tau_i|).$$

The MSE results are reported in Table 2. These results seem to indicate that our theory is applicable for a variety of models for the $\{\sigma_i\}$.

Statistical inference for β is also possible. We use the following generalization of methods suggested by Huber (1973) for the homoscedastic case. Using (2.8) of Theorem 1, we estimate the covariance of $N^{\frac{1}{2}}(\hat{\beta} - \beta)$ by

$$\text{(4.4)} \qquad \kappa(\widehat{E\psi^2})\widehat{S}^{-1}(\widehat{E\psi'})^{-2},$$

where

$$\lambda = \widehat{E\psi'} = N^{-1}\Sigma\psi'((Y_i - x_i\hat{\beta})/\hat{\sigma}_i)$$

$$K = 1 + (p + 2)(1 - \lambda)/(N\lambda)$$

$$\hat{S} = N^{-1}\Sigma x_i x_i' \hat{\sigma}_i^{-2}$$

and $\widehat{E\psi^2}$ defined similarly to $\widehat{E\psi'}$. In our Monte-Carlo experiment we constructed confidence intervals for the slope parameter β_1 in (4.1), using (4.4) and t-percentage points with $N - p - r = N - 4$ degrees of freedom. The intended coverage probability was 95%; in none of the cases did the achieved coverage probability fall below 92 , and in the majority of the cases it was at least 9[illegible].

We also attempted to solve equations (2.4) and (3.4) simultaneously using the IMSL routine ZSYSTM. Our experience was much like that of Frochlich (1973) in that the algorithm converged most of the time but not always. Dent and Hildreth (1977) were able to show that the difficulties experienced by Froehlich could be overcome by sophisticated optimization techniques. We suspect that the same holds for our problem.

The particular method for estimating $\theta = (\alpha, \theta_*)$ outlined in Section 3 and explored in this section is recommended for models such as (1.3a) - (1.3c) which satisfy (2.9). In model (1.3d) an alternative procedure is preferable, because we can exploit the relationship

$$\sigma_i^2 = \alpha_1 + \alpha_2\tau_i^2.$$

Here one would obtain initial estimates of (α_1, α_2) by (robust) regression techniques, as along the lines of Jobson and Fuller (1980), working with the squares of the residuals from a preliminary fit. One would then do one-step of a Newton-Raphson towards solving versions of (3.4) which are obtained by working with (α_1, α_2) and following the line of reasoning in (3.1) - (3.4). Monte-Carlo work

which will be reported elsewhere indicates that this technique can be quite successful.

5. Feedback. In the case of normal errors, Jobson and Fuller have suggested using the information about β in the terms $\sigma_i = H(\tau_i, \theta)$. This essentially reduces to maximizing (3.1) jointly in (θ, β). In a very nice result they show that if the error distribution is exactly normal and if (5.1) is exactly correct, then improvement over the weighted least squares estimate can be achieved. It is clear that such feedback procedures will be adversely affected by outliers or non-normal error distributions, and it is not clear how to robustly modify them.

In cases where using feedback is comtemplated, a second form of robustness must also be considered, i.e., robustness against misspecification of the function H in (3.1). Carroll (1981, unpublished) has shown that as long as H is correctly specified to order $O(N^{-1/2})$, the asymptotic properties of the weighted estimates ((2.4), (3.5)) are the same as if H were correctly specified; in this sense our weighted estimates are robust against small errors in specifying H. Carroll also shows that such robustness is not the case for feedback estimates. In fact, any gain from feedback can be more than offset by slight errors in specifying H. Since our primary interest is in β and $\sigma_i = H(\tau_i, \theta)$ is at best an approximation, we suggest that feedback should not be automatically preferred in practical use.

6. An example. In Figure #1 we plot the outcomes of 113 observations of Total Esterase (C_i) and Radioimmunoassay -RIA (Y_i), made available to us by Drs. D. Horowitz and D. Proud of the National Heart, Lung and Blood Institute. The data are clearly heteroscedastic, so we fit the model (4.1) with variance model (1.3a), estimation done as in the previous section. The results are summarized in Table #3. Since θ appears to be fairly large, the results of the Monte-Carlo indicate that weighting should be of real benefit. The confidence limits on $\hat{\theta}$

were obtained by bootstrapping (using 60 simulations). In the weighted cases the standard errors for β_0 and β_1 were obtained from (4.4), as well as by the bootstrap. The weighted results are fairly close together. While our purpose in presenting the numbers is merely illustrative, we note that the values of θ suggest that a logarithmic or square root transformation might stabilize the variances (see Box and Hill (1974)). We also fit a quadratic model to the data with little change.

REFERENCES

Anscombe, F. J. (1961). Examination of residuals. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* (J. Neyman, editor), pp 1-36. University of California Press, Berkeley and Los Angeles, California.

Bickel, P. J. (1975). One-step Huber estimates in the linear model. *J. Am. Statist. Assoc.* 70, pp 428-434.

Bickel, P. J. (1978). Using residuals robustly I: Tests for heteroscedasticity, nonlinearity. *Ann. Statist.* 6, pp 266-291.

Bickel, P. J. and Doksum, K. A. (1981). An analysis of transformations revisited. To appear in *J. Am. Statist. Assoc.*

Box, E. P. and Hill, J. (1974). Correcting inhomogeneity of variance with power transformation weighting. *Technometrics* 16, pp 385-389.

Carroll, R. J. (1980). A robust method for testing transformations to achieve approximate normality. *J. Royal Statist. Soc.*, Series B, pp 71-78.

Carroll, R. J. and Ruppert, D. (1981). On robust tests for heteroscedasticity. To appear in *Ann. Statist.*

Dent, W. T. and Hildreth, C. (1977). Maximum likelihood estimation in random coefficient models. *J. Am. Statist. Assoc.* 72, pp 69-72.

Froehlich, B. R. (1973). Some estimators for a random coefficient regression model. *J. Am. Statist. Assoc.* 68, pp 329-335.

Fuller, A. and Rao, J. N. K. (1978). Estimation for a linear regression model with unknown diagonal covariance matrix. *Ann. Statist.* 6, pp 1149-1158.

Huber, J. (1973). Robust regression: Asymptotics, conjectures and Monte-Carlo. *Ann. Statist.* 5, pp 799-821.

Huber, P. J. (1977). *Robust Statistical Procedures.* SIAM, Philadelphia.

Jobson, J. D. and Fuller, W. A. (1980). Least squares estimation when the covariance matrix and parameter vector are functionally related. *J. Am. Statist. Assoc.* 75, pp 176-181.

Jurecková, (1977). Asymptotic relations of M-estimates and R-estimates in linear regression model. *Ann. Statist.* 5, pp 464-472.

Krasker, W. S. and Welsch, R. E. (1981). Efficient bounded-influence regression estimation using alternative definitions of sensitivity. To appear in *J. Am. Statist. Assoc.*

Table #1

Ratio of Monte-Carlo MSE for the different estimators to the Monte-Carlo MSE of the "optimal" robust estimate with known weights. These MSE's are for estimating the slope in the model $Y_i = \beta_0 + \beta_1 c_i + \sigma_i \varepsilon_i$, where σ_i satisfies (1.3a).

	Sample Size N = 21 $\beta_0 = 2.0$, $\beta_1 = 1.0$ Normal Errors			Sample Size N = 41 $\beta_0 = 4.0$, $\beta_1 = 2.0$ Contaminated Errors		
Estimator	$\theta = 0.0$	$\theta = .5$	$\theta = 1.0$	$\theta = 0.0$	$\theta = .5$	$\theta = 1.0$
Unweighted LSE	.98	1.18	1.67	1.24	1.51	2.31
"Optimal" WLSE, known weights	.98	.98	.98	1.24	1.19	1.18
Our WLSE, estimated weights	1.14	1.13	1.11	1.29	1.25	1.26
Unweighted robust estimate	1.00	1.18	1.66	1.00	1.21	1.79
Our weighted robust estimate, estimated weights	1.14	1.13	1.10	1.03	1.04	1.07

Table #2

Ratio of Monte-Carlo MSE for the different estimators to the Monte-Carlo MSE of the "optimal" robust estimate with known weights. These MSE's are for estimating the slope in the model $Y_i = \beta_0 + \beta_1 c_i + \sigma_i \epsilon_i$, where σ_i satisfies $\sigma_i = \exp(\theta|c_i|)$.

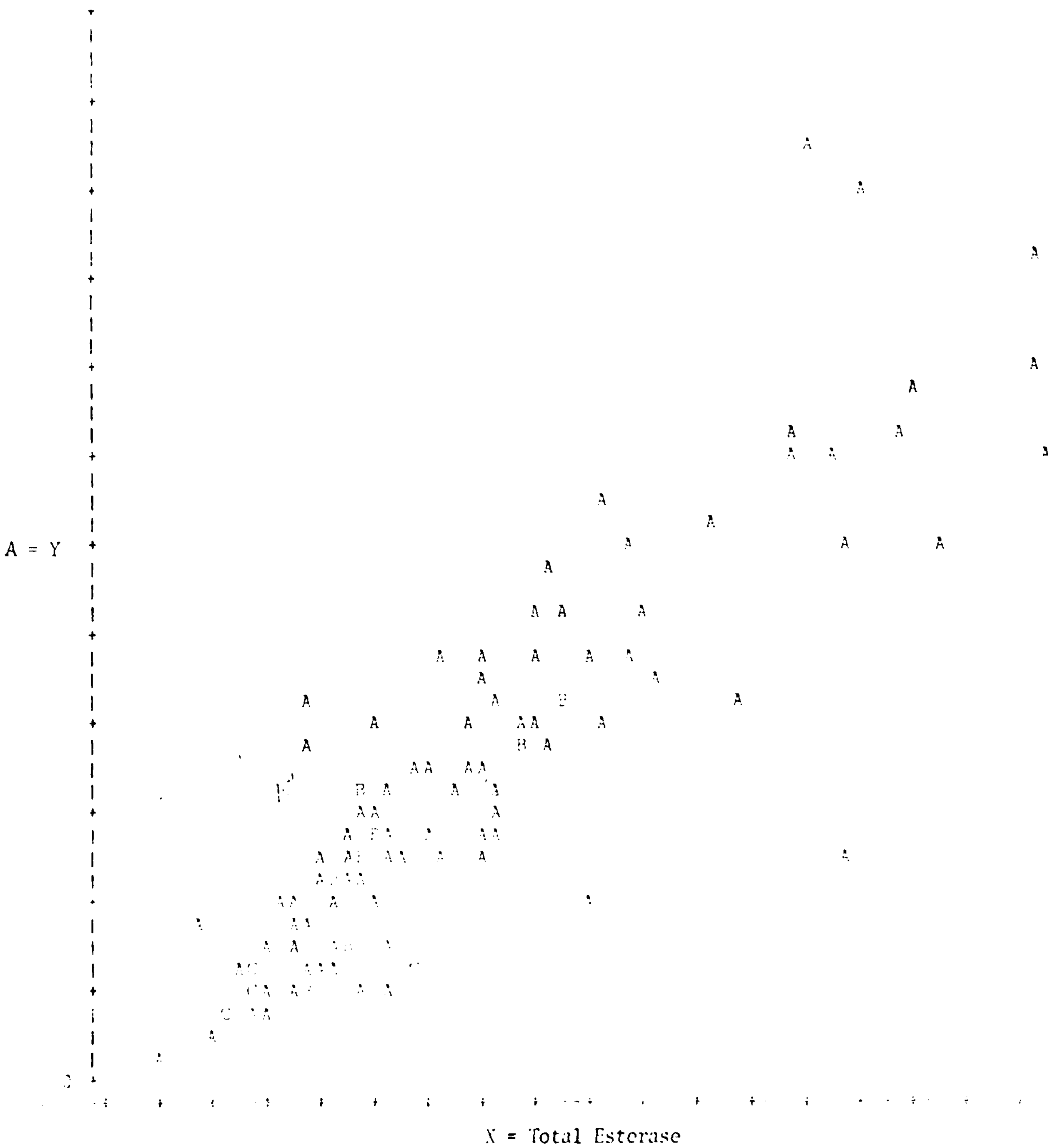
Estimator	Sample Size N = 21, $\beta_0 = 2.0$, $\beta_1 = 1.0$, Normal Errors		Sample Size N = 41, $\beta_0 = 4.0$, $\beta_1 = 2.0$, Contaminated Errors	
	$\theta = 0.0$	$\theta = .25$	$\theta = 0.0$	$\theta = .25$
Unweighted LSE	.98	1.31	1.24	2.55
"Optimal" WLSE, known weights	.98	.98	1.24	1.19
Our WLSE, estimated weights	1.15	1.13	1.28	1.26
Unweighted robust estimate	1.00	1.31	1.00	1.85
Our weighted robust estimate, estimated weights	1.14	1.13	1.03	1.05

Table #3

Results of the analysis on the data for Figure #1.

Method	$\hat{\beta}_0$	Standard Error	$\hat{\beta}_1$	Standard Error	$\hat{\theta}$	90% Confidence Limit
Unweighted least squares	-6.30	20.0	16.73	.9		
Our weighted least squares	-19.22	14.1	17.42	.9	.68	(.4,.9)
Unweighted robust	-6.54	17.4	16.67	.8		
Our weighted robust	-26.99	11.8	17.73	.9	.85	(.7,1.1)

FIGURE #1



APPENDIX A

The following general theorem will be used when studying $\hat{\underline{\beta}}_0$, $\hat{\underline{\theta}}$ and $\hat{\underline{\beta}}$.

<u>Theorem A.1</u>. Let g_i, k_i, and $A(\delta,i)$ $(= g_{iN}$, k_{iN}, and $A(\delta,i,N))$ be sequences of positive constants such that

$$\text{(A.1)} \qquad \lim_{N\to\infty}\Big(\sup_{i\le N} k_i + k_i g_i + N^{-1/2} A(\delta,i)\Big) = 0$$

and

$$\text{(A.2)} \qquad \sup_{N}\Big(\sum_{i=1}^{N} k_i^2 + k_i^2 g_i^2 + N^{-1/2} g_i k_i\Big) = C_1 < \infty \ .$$

Let ϕ_i be a function from R^3 to R^1 satisfying

$$\text{(A.3)} \qquad E\phi_i(\varepsilon_1,0,0) = 0 \text{ for all } i,$$

$$\text{(A.4)} \qquad \lim_{k\to 0} \sup E\{\sup|\phi_i(\varepsilon_1,r,s) - \phi_i(\varepsilon_1,r',s')|: |r|,|r'|,|s|,|s'| \le k \text{ and } |r - r'|,|s - s'| \le k\delta\} \le C_0 \delta g_i$$

for some C_0 and all $0 < \delta \le 1$ and all i,

$$\text{(A.5)} \qquad \sup_{i\le N} g_i^{-1}\, E(\phi_i(\varepsilon_1,r,s) - \phi_i(\varepsilon_1,0,0) - A(\delta,i)r) = o(|r| + |s|) \ ,$$

$$\text{(A.6)} \qquad \lim_{r,s\to 0}\ \sup_{i\le N} g_i^{-2}\, E(\phi_i(\varepsilon_1,r,s) - \phi_i(\varepsilon_1,0,0))^2 = 0 \ ,$$

and $\lim\limits_{N\to\infty} \sup\limits_{i\le N} g_i^{-2}\, E\phi_i^2(\varepsilon_i,0,0) < \infty$. Let $\alpha_i^{(1)}$, $\alpha_i^{(2)}$, and $\underline{\alpha}_i^{(3)}$ be functions from

R^m to R^1, R^1 and R^n respectively, and let $z_1, \ldots, z_N$ be elements of R^n satisfying

$$\text{(A.7)} \qquad \alpha_i^{(j)}(0) = 0 \ ,$$

and

$$\text{(A.8)} \qquad ||\alpha_i^{(j)}(x) - \alpha_i^{(j)}(y)|| \le k_i ||x - y|| \ \text{for } j = 1,2$$

and $||\alpha_i^{(3)}(x) - \alpha_i^{(3)}(y)|| \le k_i$ [illegible] $||x - y||$ [illegible] all x and y in R^m, $i = 1, \ldots, N$, and

$$\text{(A.9)} \qquad [illegible] ||z_i|| \le k_i \ .$$

For $\Delta \in R^m$, define the process

$$U_N(\Delta) = N^{-\frac{1}{2}} \sum_{i=1}^{N} [illegible] \alpha_i^{(1)}(\Delta) [illegible] \alpha_i^{(2)}(\Delta)) (z_i + \alpha_i^{(3)}(\Delta)) \ .$$

Then, for all $M > 0$,

$$\text{(A.10)} \qquad \sup_{||\Delta|| \le M} ||U_N(\Delta) - U_N(0) - N^{-\frac{1}{2}} \sum_{i=1}^{N} \lambda(0,i) \alpha_i^{(1)}(\Delta) z_i|| = o_p(1) \ .$$

Proof of Theorem A.1. Fix $M > 0$. We will show that

$$\text{(A.11)} \qquad U_N(\Delta) - U_N(0) = N^{-\frac{1}{2}} \sum_{i=1}^{N} \lambda(0,i) \alpha_i^{(1)}(\Delta) z_i + o_p(1)$$

and

(A.12) $\qquad U_N(\underline{\Delta}) - U_N(\underline{0}) - E(U_N(\underline{\Delta}) - U_N(\underline{0})) = o_p(1)$

for each fixed Δ, and that there exists K depending upon M but not δ such that, for all $0 < \delta < 1$ and all N,

(A.13) $\qquad E \sup\{||U_N(\underline{\Delta}) - U_N(\underline{\Delta}^*)||: ||\underline{\Delta}|| \le M,\ ||\underline{\Delta}^*|| \le M,\ ||\underline{\Delta} - \underline{\Delta}^*|| \le M\delta\} < K\delta$.

Since for any δ, we can cover a ball of radius M in R^m with a finite number of balls of radius δ, (A.11), (A.12) and (A.13) prove that, for each $0 < \delta < 1$,

$$\lim_{N\to\infty} P\Big(\sup_{||\underline{\Delta}||\le M} ||U_N(\underline{\Delta}) - U_N(\underline{0}) - N^{-1/2} \sum_{i=1}^{N} A(\phi,i)\alpha_i^{(1)}(\underline{\Delta})\underline{z}_i|| < K\delta\Big) = 1 ,$$

which proves the theorem.

To prove (A.11), note that by (A.3),

$$N^{-1/2} \sum_{i=1}^{N} E(U_N(\underline{\Delta}) - U_N(\underline{0}))$$

$$= N^{-1/2} \sum_{i=1}^{N} E(\phi_i(\varepsilon_i, \alpha_i^{(1)}(\underline{\Delta}), \alpha_i^{(2)}(\underline{\Delta})) - \phi_i(\varepsilon_i,0,0))(\underline{z}_i + \underline{\alpha}_i^{(3)}(\underline{\Delta})) .$$

We next have by (A.1), (A.7) and (A.8) that for all large N,

(A.14) $\qquad ||\underline{z}_i + \underline{\alpha}_i^{(3)}(\underline{\Delta})|| \le 2||\underline{z}_i||$,

and also by (A.5), (A.7) and (A.8),

$$E(\phi_i(\varepsilon_i, \alpha_i^{(1)}(\underline{\Delta}), \alpha_i^{(2)}(\underline{\Delta})) - \phi_i(\varepsilon_i,0,0)) = A(\phi,i)\alpha_i^{(1)}(\underline{\Delta}) + o(g_i k_i ||\underline{\Delta}||)$$

uniformly in i. Therefore,

$$E((U_N(\underline{\Delta}) - U_N(\underline{0})) = N^{-\frac{1}{2}} \sum_{i=1}^{N} A(\phi,i)\alpha_i^{(1)}(\underline{\Delta})\underline{z}_i + o(N^{-\frac{1}{2}} \sum_{i=1}^{N} g_i k_i ||\underline{z}_i||$$

$$+ N^{-\frac{1}{2}} \sum_{i=1}^{N} A(\phi,i)\alpha_i^{(1)}(\underline{0})\alpha_i^{(2)}(\underline{\Delta}))\ .$$

By (A.1), (A.7) and (A.8), the last term on the RHS is $o(1)$. By (A.2), (A.9) and the Cauchy-Schwarz inequality, the second term is bounded by

$$o[\sum_{i=1}^{N} g_i k_i^2] = o(1)$$

so that (A.11) holds. Then, again using (A.14), we have that for N large,

$$\tfrac{1}{2}\,\mathrm{Var}(U_N(\underline{\Delta}) - U_N(\underline{0}))$$

$$\le (2N^{-1} \sum_{i=1}^{N} g_i^2 ||\underline{z}_i||^2) \sup_{i\le N} g_i^{-2}\, E(\phi_i(\varepsilon_i\,,\alpha_i^{(1)}(\underline{\Delta})\,,\alpha_i^{(2)}(\underline{\Delta}))$$

$$- \phi_i(\varepsilon_i,0,0))^2 + E||N^{-\frac{1}{2}} \sum_{i=1}^{N} \phi_i(\varepsilon_i,0,0)\underline{\alpha}_i^{(3)}(\underline{\Delta})||^2\ .$$

The second term on the RHS is $o(1)$ by (A.6)-(A.8). It also follows from (A.6)-(A.8) that

$$\sup_{i\le N} g_i^{-2}\, E(\phi_i(\varepsilon_i\,,\alpha_i^{(1)}(\underline{\Delta})\,,\alpha_i^{(2)}(\underline{\Delta})) - \phi_i(\varepsilon_i,0,0))^2 = o(1)\ .$$

Therefore, (A.12) is proved by applying (A.2) and (A.9). Finally, by (A.14) the RHS of (A.13) is less than or equal to

$$2N^{-\frac{1}{2}} \sum_{i=1}^{N} ||\underline{x}_i|| E \sup\{|\phi_i(\varepsilon_i, \alpha_i^{(1)}(\underline{\Delta}), \alpha_i^{(2)}(\underline{\Delta}))$$

$$- \phi_i(\varepsilon_i, \alpha_i^{(1)}(\underline{\Delta}^*), \alpha_i^{(2)}(\underline{\Delta}^*)|: ||\underline{\Delta}|| \le M, ||\underline{\Delta}^*|| \le M,$$

$$||\underline{\Delta} - \underline{\Delta}^*|| \le M\delta\}$$

$$+ N^{-\frac{1}{2}} \sum_{i=1}^{N} E \sup\{\phi_i(\varepsilon_i, \alpha_i^{(1)}(\underline{\Delta}^*), \alpha_i^{(2)}(\underline{\Delta}^*))(\underline{\alpha}_i^{(3)}(\underline{\Delta}) - \underline{\alpha}_i^{(3)}(\underline{\Delta}^*)):$$

$$||\underline{\Delta}|| \le M, ||\underline{\Delta}^*|| \le M, ||\underline{\Delta} - \underline{\Delta}^*|| \le M\delta\} .$$

By (A.2), (A.8), (A.9) and the Cauchy-Schwarz inequality, the first term does not exceed

$$\sup_N \left(\sum_{i=1}^{N} (g_i k_i^2)\right) C_0 \delta = C_0 C_1 \delta .$$

By (A.2), (A.4) and (A.9), the second term does not exceed

$$\delta M^2 C_0 N^{-\frac{1}{2}} \sum_{i=1}^{N} g_i k_i^2 (1 + ||\underline{z}_i||) = o(1) .$$

Therefore, (A.13) is verified. $\square$

APPENDIX B

Proof of Theorem 1. For $\underline{\Delta}_1$ and $\underline{\Delta}_3$ in R^p, $\underline{\Delta}_2$ in R^r, and $\underline{\Delta} = (\underline{\Delta}_1, \underline{\Delta}_2, \underline{\Delta}_3)$, define

$$\alpha_i^{(1)}(\underline{\Delta}) = N^{-\frac{1}{2}} \underline{d}_i \underline{\Delta}_1$$

$$h_i(\underline{\Delta}) = h(t_i + x_i \Delta_3 N^{-\frac{1}{2}})$$

$$\alpha_i^{(2)}(\underline{\Delta}) = \exp(-h_i(\underline{\Delta})'\underline{\Delta}_2 N^{-\frac{1}{2}}) + (h_i(\underline{0}) - h_i(\underline{\Delta}))\theta - 1$$

and

$$\alpha_i^{(3)}(\underline{\Delta}) = d_i \alpha_i^{(2)}(\underline{\Delta}) \ .$$

Define the process

$$U_N(\underline{\Delta}) = N^{-\frac{1}{2}} \sum_{i=1}^{N} \psi[(\epsilon_i - \alpha_i^{(1)}(\underline{\Delta}))(1 + \alpha_i^{(2)}(\underline{\Delta}))](d_i + \alpha_i^{(3)}(\underline{\Delta})) \ .$$

Note that (2.4) can be rewritten as

(B.1) $$U_N(N^{\frac{1}{2}}(\hat{\beta} - \beta), N^{\frac{1}{2}}(\hat{\theta} - \theta), N^{\frac{1}{2}}(\hat{\sigma}_0 - \sigma)) = 0 \ .$$

Letting $g_i \equiv 1$, $k_i = N^{-\frac{1}{2}}(1 + ||\underline{d}_i|| + ||h(t_i)||)$, $\psi_i(\epsilon_i, r, s) = \psi((\epsilon_i - r)(1 + s))$, $d_i = z_i$, and $A(\psi_i) = A(\psi) = E\psi'$, the conditions of Theorem A.1 are implied by (2.5) and B1 - B3, so for all $M > 0$,

(B.2) $$\sup_{||\underline{\Delta}|| \le M} ||U_N(\underline{\Delta}) - U_N(\underline{0}) - A(\psi) S \underline{\Delta}_1|| = o_p(1) \ .$$

Now by Chebychev's theorem, B1 and B5,

$$U_N(\underline{0}) = O_p(1) \ .$$

In proving the Theorem, we will not assume that $\hat{\beta}$ actually solves (2.4), but rather that the l.h.s. of (2.4) evaluated at $\hat{\beta}$ is less than twice it's infimum over all β. However, as seen in Huber (1977), (2.4) will have a unique solution if ψ is monotone. From the last equation, we have

$$\underline{\Delta}^* = -(A(\psi)S)^{-1}\, U_N(0) = O_p(1) \ ,$$

so that by (B.2), $U(\underline{\Delta}^*) = o_p(1)$. Consequently, by (B.1),

(B.3) $$U_N(N^{\frac{1}{2}}(\hat{\beta}-\beta)\ , N^{\frac{1}{2}}(\hat{\theta}-\theta)\ , N^{\frac{1}{2}}(\hat{\beta}_0-\beta)) = o_p(1) \ .$$

By (2.1), $(\hat{\theta}-\theta) = O_p(N^{-\frac{1}{2}})$, so we need only establish that

(B.4) $$(\hat{\beta}-\beta) = O_p(N^{-\frac{1}{2}})$$

to conclude from (B.1) and (B.2) that (2.7) holds. But by (B.3), (B.4) holds if for each $\eta > 0$, $\varepsilon > 0$ and M_1, there exists M_2 satisfying

(B.5) $$P\Big(\inf_{||\underline{\Delta}_1|| \ge M_2}\ \inf_{||\underline{\Delta}_2|| \le M_1}\ \inf_{||\underline{\Delta}_3|| \le M_1} ||U_N(\underline{\Delta})|| > \eta\Big) > 1 - \varepsilon \ .$$

Now (B.5) follows from (B.2) in a manner quite similar to Jurečková's (1977) proof of her Lemma 5.2. □

APPENDIX C

<u>Proof of Theorem 2</u>. For $\underline{\Delta}_1$ in R^p, $\underline{\Delta}_2$ in R^q, and $\underline{\Delta} = (\underline{\Delta}_1, \underline{\Delta}_2)$, define

$$h_i(\underline{\Delta}) = h(\tau_i + x_i \underline{\Delta}_1 N^{-\frac{1}{2}}) ,$$

$$\alpha_i^{(1)}(\underline{\Delta}) = \exp\{-h_i(\underline{\Delta}) \underline{\Delta}_2 N^{-\frac{1}{2}} + (h_i(\underline{0}) - h_i(\underline{\Delta}))\theta\} - 1 , \tag{C.1}$$

$$\alpha_i^{(2)}(\underline{\Delta}) = N^{-\frac{1}{2}} d_i \underline{\Delta}_1 \quad \text{(see (2.5))} , \tag{C.2}$$

and

$$\alpha_i^{(3)}(\underline{\Delta}) = h_i(\underline{0}) - h_i(\underline{\Delta}) . \tag{C.3}$$

Then let $\phi(x,y,z) = \chi((x - z)(1 + y))$ and define the process

$$W_N(\underline{\Delta}) = -N^{-\frac{1}{2}} \sum_{i=1}^{N} \phi(\varepsilon_i , \alpha_i^{(1)}(\underline{\Delta}) , \alpha_i^{(2)}(\underline{\Delta}))(h(\tau_i) + \alpha_i^{(3)}(\underline{\Delta})) .$$

Note that (3.6) can be written as

$$||W_N(N^{\frac{1}{2}}(\hat{\beta}_0 - \beta) , N^{\frac{1}{2}}(\hat{\theta} - \theta))|| = \text{minimum} .$$

However, by (3.5), <u>C1</u> and Chebyshev's inequality,

$$W_N(\underline{0}) = O_p(1) \tag{C.4}$$

so that

$$W_N(N^{\frac{1}{2}}(\hat{\beta}_0 - \beta) , N^{\frac{1}{2}}(\hat{\theta} - \theta)) = O_p(1) .$$

We can therefore prove (3.7) by showing that for each $M_1 > 0$, $\varepsilon > 0$ and $Q > 0$, there exists $M_2 > 0$ such that

(C.5) $\quad P[\inf\{||W_N(\underline{\Delta})||: ||\underline{\Delta}_1|| \le M_1,\ ||\underline{\Delta}_2|| \ge M_2\} > Q] \ge 1 - \varepsilon$.

We will prove (C.5) by modifying the proof of Jurečková's (1977) Lemma 5.2. We first apply Theorem A.1 with $z_i = h_i(\underline{0})$, $g_i \equiv 1$, $A(\rho,i) = A(X)$, and $k_i = N^{-1/2}(h(\tau_i) + ||x_i|| + ||d_i||)$. Then

$$\sup_{||\underline{\Delta}||\le M} ||W_N(\underline{\Delta}) - W_N(\underline{0}) - A(X)N^{-1/2}\sum_{i=1}^{N} h(\tau_i)\alpha_i^{(1)}(\underline{\Delta})|| = o_p(1) .$$

By a Taylor series expansion,

$$\alpha_i^{(1)}(\underline{\Delta}) = -N^{-1/2} h(\tau_i)\underline{\Delta}_2 + (h_i(\underline{0}) - h_i(\underline{\Delta}))\theta + o_p(N^{-1/2}) .$$

Thus, by $\underline{C5}$ setting

$$G_N(\underline{\Delta}) = N^{-1/2} \sum_{i=1}^{N} h(\tau_i)^T (h_i(\underline{0}) - h_i(\underline{\Delta}))\theta ,$$

we obtain

(C.6) $$\sup_{||\underline{\Delta}||\le M} ||W_N(\underline{\Delta}) - W_N(\underline{0}) - A(X)H\underline{\Delta}_2 + G_N(\underline{\Delta})|| = o_p(1) .$$

Now fix $\varepsilon > 0$, $M_1 > 0$, $Q > 0$. Use (C.3) to choose γ such that

$$P(||W_N(\underline{0})|| \ge \gamma/2) < \varepsilon/2 .$$

Define

$$D = \sup_N \sup_{||\underline{\Delta}_1||\le M_1} ||G_N(\underline{\Delta})|| .$$

Then $D < \infty$ (G_N depends only on $\underline{\Delta}_1$). Define M_2 by $[\lambda(X)\lambda_\infty M_2/2 - \gamma - D] = Q$.

Using C5 and (C.6), find N_0 such that $\lambda_N \geq \lambda_\infty/2$ and

$$P\{\sup_{||\underline{\Delta}_2||=M \text{ and } ||\underline{\Delta}_1||\leq M_1} ||W_N(\underline{\Delta}) - W_N(\underline{0}) - A(X)H\underline{\Delta}_2 - G_N(\underline{\Delta})|| < \gamma/2\}$$

$$\geq 1 - \varepsilon/2 \quad (N \geq N_0) .$$

If $||\underline{\Delta}_2|| = M_2$, $||\underline{\Delta}_1|| \leq M_1$, and $N \geq N_0$, then with probability at least $1 - \varepsilon$,

$$\begin{aligned}\underline{\Delta}_2 W_N(\underline{\Delta}) &\geq -M_2||W_N(\underline{0})|| + \underline{\Delta}_2^T H\underline{\Delta}_2 A(X) - M_2 D - M_2\gamma/2 \\ &\geq [A(X)\lambda_\infty M_2/2 - \gamma - D]M_2 = QM_2 .\end{aligned}$$

Since X is nondecreasing on $[0,\infty)$ by C1, $\underline{\Delta}_2 W_N(\underline{\Delta}_1, \underline{\Delta}_2 s)$ is a nondecreasing function of s. Thus, $||\underline{\Delta}_2|| \geq M_2$ implies

$$\begin{aligned}\underline{\Delta}_2 W_N(\underline{\Delta}) &\geq \underline{\Delta}_2 W_N(\underline{\Delta}_1, M_2\underline{\Delta}_2||\underline{\Delta}_2||^{-1}) \\ &\geq (||\underline{\Delta}_2||/M_2)(M_2\underline{\Delta}_2||\underline{\Delta}_2||^{-1} W_N(\underline{\Delta}_1, M_2\underline{\Delta}_2||\underline{\Delta}_2||^{-1})) \\ &\geq ||\underline{\Delta}_2||Q .\end{aligned}$$

Thus,

$$P(\inf_{\substack{||\underline{\Delta}_1||\leq M_1 \\ ||\underline{\Delta}_2||\geq M_2}} \frac{\underline{\Delta}_2 W_N(\underline{\Delta})}{||\underline{\Delta}_2||} \geq Q) \geq 1 - \varepsilon ,$$

which with the Cauchy-Schwarz inequality proves (3.2). □

UNCLASSIFIED

18 REPORT DOCUMENTATION PAGE

19 AFOSR-TR-81-0219

6 Robust Estimation in Heteroscedast Linear Models

9 Interim rept.

10 Raymond J. Carroll ~~and~~ David Ruppert

15 AFOSR-80-0080, VNSF-MCS78-01240

University of North Carolina/Chapel Hill
Department of Statistics
Chapel Hill, N. C. 27514

61102F 16 2304 A5 17 A5

Air Force Office of Scientific Research
Bolling Air Force Base
Washington, DC 20332

11 Feb[redacted]81

32

12 34

UNCLASSIFIED

16. DISTRIBUTION STATEMENT

Approved for Public Release -- Distribution Unlimited

17. DISTRIBUTION STATEMENT

18. SUPPLEMENTARY NOTES

19. KEY WORDS

weighted M-estimates, weighted least squares, unknown variances, estimated weights, asymptotically normal

20. ABSTRACT

We consider a heteroscedastic linear model in which the variances are a parametric function of the mean responses $(\mu_1, \ldots, \mu_N)$ and a parameter θ. We propose robust estimates for the regression parameter β and show that, as long as a reasonable starting estimate of θ is available, our estimates of β are asymptotically equivalent to the natural estimate obtained with known

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

182850

20. variances. A particular method for estimating θ is proposed and shown by Monte-Carlo to work quite well, especially in power and exponential models for the variances. We also briefly discuss a "feedback" estimate of β.

END

DATE